Edge Computing with Deep Learning and Internet of Things for Recognising and Predicting Students Emotions and Mental Health

[1,2]Shaymaa Hussein Nowfal, [3]Firas Tayseer Ayasrah, [4]Vijaya Bhaskar Sadu, [5]Jasmine Sowmya V, [6]Subbalakshmi A V V S and [7]Kamal Poon

[1]Medical Physics Department, College of Science, University of Warith Al-Anbiyaa, Karbala, Iraq.
[2]Medical Physics Department, College of Applied Medical Sciences, University of Kerbala, Karbala, Iraq.
[3]College of Education, Humanities and Science, Al Ain University, Al Ain, Abu Dhabi, United Arab Emirates.
[4]Department of Mechanical Engineering, Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India.
[5]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.
[6]Department of Commerce, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore, Tamil Nadu India.
[7]College of Science and Engineering, Southern Arkansas University, Magnolia, AR, 71753, USA.
[1,2] shaymaa@uowa.edu.iq, [3]firas.ayasrah@aau.ac.ae, [4]sadhu.vijay@gmail.com, [5]vemulajasmine@kluniversity.in, [6]subbusravani76@gmail.com, [7]kamalpoon57@gmail.com

Correspondence should be addressed to Firas Tayseer Ayasrah: firas.ayasrah@aau.ac.ae

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202404101
Received 31 May 2024; Revised from 02 August 2024; Accepted 10 August 2024
Available online 05 October 2024.

Abstract – A person's Mental Health (MH) dramatically influences their complete evolution in life, including their cognitive, emotional, and psychomotor components. A person with good MH is content with life and can be creative, learn new things, and take risks to accomplish more significant objectives. Currently, college students are dealing with MH concerns for various causes, which affect their academic performance and significantly contribute to poor academic results. Therefore, encouraging MH in college students presents a significant problem for educators, parents, teacher educators, and governments. Adolescence is a crucial and delicate time characterized by considerable physical, emotional, social, and religious changes. The physical, social, and psychological facets of an individual's growth are laid out in this period, with mental health as a crucial factor in promoting these gains. Therefore, it becomes crucial for researchers to use Deep Learning (DL) algorithms to study the association between MH and vital psychological characteristics, such as emotional intelligence, personality traits, and intelligence. The personal aspects, namely personality, emotional intelligence, and MH are all related ideas that influence one another. Individuals must have mental well-being and emotional harmony to have a good personality. The current study uses DL techniques to investigate the relationship between college students' MH, emotional intelligence, and personality features. To perform a thorough study on emotion identification and Mental Health Prediction (MHP) among college students, this project investigates the integration of edge computing enabled by the Internet of Things (IoT) in the context of intelligent systems. Innovative treatments are urgently needed due to this population's rising prevalence of MH issues. This paper aims to continuously monitor and predict college students' MH using Edge Computing (EC) and IoT technology.

Keywords – Mental Health, Facial Expression, Higher Education, Emotion, Deep Learning, and Accuracy.

I. INTRODUCTION

Nowadays, computing systems are united with humans' day-to-day lives and functions as a reflection of humans. To achieve this goal entirely, it is necessary to build human-based user interfaces that quickly respond to multi-modal human-machine communication systems by next-generation computing systems [1]. The interfaces are designed so that they can recognize and understand the emotions, meanings, and intentions communicated by affective and social variable signals. The automatic analysis of non-verbal behaviour, particularly facial behaviour, Plays a significant role in digital image processing, pattern identification, Human-Computer Interaction (HCI), and Computer Vision (CV). Facial Expressions

(FE) are some generalized way for human to express their emotions to others. The FE is analyzed effectively using initiative-taking and effective user interfaces, patient-profiled wellness technologies, and learner-adaptive tutoring models.

The FE identification method is widely used in social networking, human-computer interfaces, real-time security surveillance systems, and subject-tracking applications. Several methods are developed for identifying FE, and in this approach, FE, like sadness, anger, happiness, surprise, fear, and disgust are identified. The traditional FE identification approaches are designed to enhance facial features' identification rate using static images.

Nevertheless, this new method provides a robust implementation of FE identification from vast image databases, which changes according to time and space. Thus, it is essential to design a model that can be utilized in personal identification systems or security systems, which include Quantitative face evaluation.

Nowadays, FE is considered the key factor in social interactions with other humans and is the best way to express emotions. The results show that FE represents 55% of the communicated message, and voice and language represent 38% and 7%, respectively. As artificial intelligence technology is developing rapidly, FE is considered an essential factor of advanced HCI. This approach aims to convey the information reliably, efficiently, and automatically. Generally, the Feature Extraction (FE) process extracts more features, but only fewer features must be chosen based on specific criteria. Also, the k-nearest Neighbour (k-NN) algorithm effectively identifies FE. Here, the spatial and frequency properties are varied based on specific vision needs. Gabor filter is utilized to evaluate facial expressions. This approach is simple, dependable, and has a high recognition rate. The performance of this method is enhanced when applied to an automatic Facial Expression Recognition (FER) system.

The traditional methods of FER are complicated, time-consuming, and labor-intensive. Therefore, the researchers established automatic FER and evaluation approaches to allow rigorous, feasible, quantitative analysis of FE in wide applications. In this innovative approach, the variations in the FE are identified and observed automatically in a digital image sequence, *i.e.,* 25 images/second [2]. The intensities of Facial Action Coding System (FACS) action units in every video image are calculated by an Artificial Neural Network (ANN)-based classifier using the FE. At the training level, the features in the sample image are retrieved, and the classifier is used to classify the features. The human faces are retrieved from the acquired face image at the testing level. At last, the identified expression of the face image is outputted.

Human FE is used in the biometric field and business, managerial, organizational, cultural contexts, Medical imaging, HCI, and Telecommunication fields. The HCI system is the one in which the computer can identify, communicate, and respond to the actions of the user based on the emotional state of the face of the human. Therefore, users will be allowed to communicate effectively. The human FE is formed by moving the facial muscles that bring FE (**Fig 1**), like raising eyebrows, compressing eyebrows, stretching lips, and opening eyes.

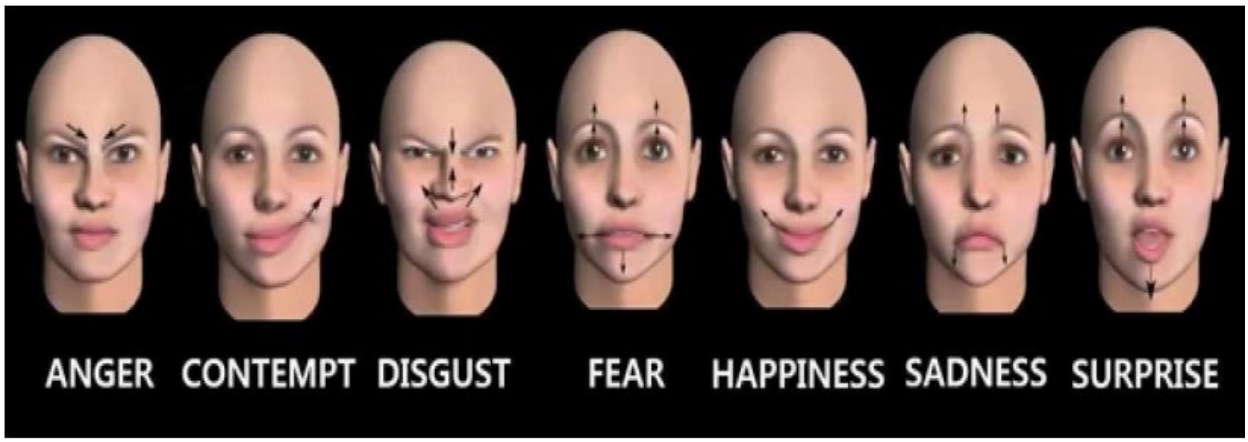

Fig 1. Facial Expression.

The human face has several hidden features for detecting the person's age, identity, and gender. This novel approach is used to detect the age of humans. This approach primarily identifies the person's age so that small, aged people's Internet use can be limited. Also, children are not permitted to access numerous sites. There are several vending machines for getting items such as alcohol cigarettes, which are not accessible to children. The gender and age of the person are recognized by visual observation of images, whereas it is difficult in computer vision. The fact is identified using nose length, lip length, and eye size. Gender is identified by determining features like Mustache, eye size, and total skin color pixels. The age is determined using the right and left cheeks, eyelid region, and forehead portion.

Motivation

Due to the growing Mental Health (MH) problem on college campuses, research on Emotion Recognition (ER) and Mental Health Prediction (MHP) in college students utilizing Deep Learning (DL) algorithms is of utmost importance. Significant pressures are frequently a part of college life, and early identification of MH difficulties by DL can enable prompt intervention and individualized treatment, potentially lowering dropout rates and encouraging secret access to MH

resources. Additionally, this study can contribute to this research's understanding of the interaction between emotions and MH by providing valuable data-driven insights into the variables causing MH problems among students and supporting evidence-based policies and programs. Furthering privacy, permission, and algorithmic bias issues, the ethical implications of using Artificial Intelligence (AI) for mental health assessment and support might be investigated. Ultimately, this research can potentially have a significant social impact by improving mental health, lowering stigma, and easing MH problems in the public health system.

Research Objective

Human-Computer Interaction (HCI) technology is measured as technological advancements that equip computers with an interface to realize the interaction between computers and humans. With faster development in recognizing patterns and AI, enormous investigations have been conducted in the technological field of HCI. FER is a significant method for intelligent HCI that has extensive background applications. It is utilized in multiple distance education, medicine, public security, and games. FER hauls out information that specifies features of FE from input images via processing technology and classifies featured FE based on emotional expressions like surprise, aversion, neutrality, and happiness. FER plays a vital role in emotional quantification.

Multiple complex problems are solved efficiently with the vast advancements in DL approaches. Here, a novel Deep Convolutional Neural Network (D-CNN) model is adopted to identify the FE with reduced complexity.

The Significance of This Research Work is Listed Below

(a) Initially, data acquisition is done by obtaining the data from the online available FER-2013 (*https://www.kaggle.com/deadskull7/fer2013*) [3].
(b) Some pre-processing steps like normalization, equalization, and bagging approaches are performed to convert the 3D/2D images to 1D images.
(c) After pre-processing, FE is done with crafter features to enhance the quality of classification outcomes or predictions.
(d) Finally, classification is performed with the D-CNN model to predict the FE with the most minor complexity.
(e) The classifier performance is measured with accuracy, sensitivity, specificity, and MCC.

To thoroughly examine emotion identification and MHP among college students, this research accomplished a complete analysis of the Internet of Things (IoT)-)-enabled Edge Computing (EC), especially within the framework of smart systems. By exploring the relationship between technology and well-being, we hope to improve our comprehension of college students' emotional experiences and offer proactive help, all while considering the ethical, privacy, and accuracy ramifications that this integration implies. This study has the potential to advance the use of EC and IoT in healthcare and education and significantly improve the MH and overall well-being of college students.

II. RELATED WORKS

A developing and essential area of study in the science of pattern recognition is Facial Emotion Recognition (FER). Nonverbal communication is crucial to human daily interactions and contributes considerably, making up between 55% and 93% of all communication. Numerous fields, including surveillance videos, expression analysis, gesture recognition, smart homes, computer games, depression treatment, patient monitoring, anxiety assessment, lie detection, paralinguistic communication, operator fatigue detection, robotics, and surveillance videos, use facial emotion analysis effectively [4]. This research thoroughly analyzes FER, utilizing literature from reliable sources released during the last ten years. Both traditional Machine Learning (ML) and DL techniques are included in the review [5]. We also examine various publicly accessible FER datasets utilized for evaluation measures and compare them to benchmark results. For aspiring researchers, this thorough analysis acts as a road map by highlighting the current inadequacies in this subject. Ultimately, this study is a great resource that helps inexperienced researchers understand the core ideas and innovative techniques in FER while giving seasoned researchers an understanding of potential directions for future research projects [6].

[7] implemented a new process of facial image identification by incorporating an improved Cat Swarm Optimization (ICSO) algorithm. The input given to the system extracts the duplicate images from the dataset and recognizes the person's emotional condition through the expressions of the face. The intense attributes in the image are acquired using the Deep Convolution Neural Networks (DCNN) method. The optimal attributes that differentiate the individual's unique expression are chosen from the image of the face using ICSO. The recovering efficiency of the system is highly enhanced by integrating DCNN and ICSO. Different expressions like happiness, sadness, anger, fear, and surprise are classified by incorporating NN and SVM using ensemble classifiers. This approach is implemented in JAFFE, Pie, and CK+ databases, and the advantages include improved efficiency, high accuracy, and less calculation time.

Design a method integrating light boost Decision Trees (DT) and Neural Networks ((NN) to attain high precision [8-10]. The DT can pull out highly differential feature vectors like local binary features for every FE across many facial points. The binary features are combined in a form to optimize FER via NN. These combined optimization techniques can highly increase the recognition rates of harder FEs like sadness and fear. This Local Binary Feature-Shallow Neural Networks (LBF-SNN) method attains vast improvement in execution time compared with previously existing algorithms.

[10-11] used a multi-label Feature Selection (FS) technique to decrease the complexity in computation, increase the interpretability of the system, and enhance classification performance by choosing a discriminative subset of attributes from high dimensional data. Correlation-based Feature Selection (CFS) is used to evaluate the relationship between the labels and features, which is integrated with Genetic Algorithm (GA) and hill-climbing techniques for performing the multi-label FS process. The existing binary bat algorithm is mutated to a newer version to arrange the components in constant size. The multi-label FS technique is implemented by maximizing the CFS criteria for selecting a constant number of discriminative features. The performance of this method is enhanced when compared with other methods concerning sample-based performance calculation.

[12-13] use filtering and edge detection methods and remove non-essential information and background noise in the pre-processing phase. The edge detectors emphasize common facial elements, identify sharp discontinuities, and erase unnecessary information. The Differences between Gaussian, Canny edge detectors, Kirsch, Sobel Prewitt, Laplacian of Gaussian, and Robert's detectors are used to pre-process face images. Local FE algorithm, Viola-Jones Facial Detection Algorithm, local directional patterns, and k-NN are adopted for image detection, FE, and classification in respective order [44-47].

[14-15] focuses on detecting an optimal subset of certain features. It used an improved Binary Bat Algorithm with a Cross-Entropy (BBAE) approach to stabilize between utilization and investigation on an evolution basis. Ten uniform benchmark datasets from the UCI repository can calculate this method's efficiency compared with wrapper FS methods like ALO, GA, and PSO. The results show the efficiency in choosing the most important classification features, so this approach effectively enhances classification accuracy.

III. PROPOSED METHODOLOGY – ER AND MHP OF COLLEGE STUDENTS BASED ON DL ALGORITHM

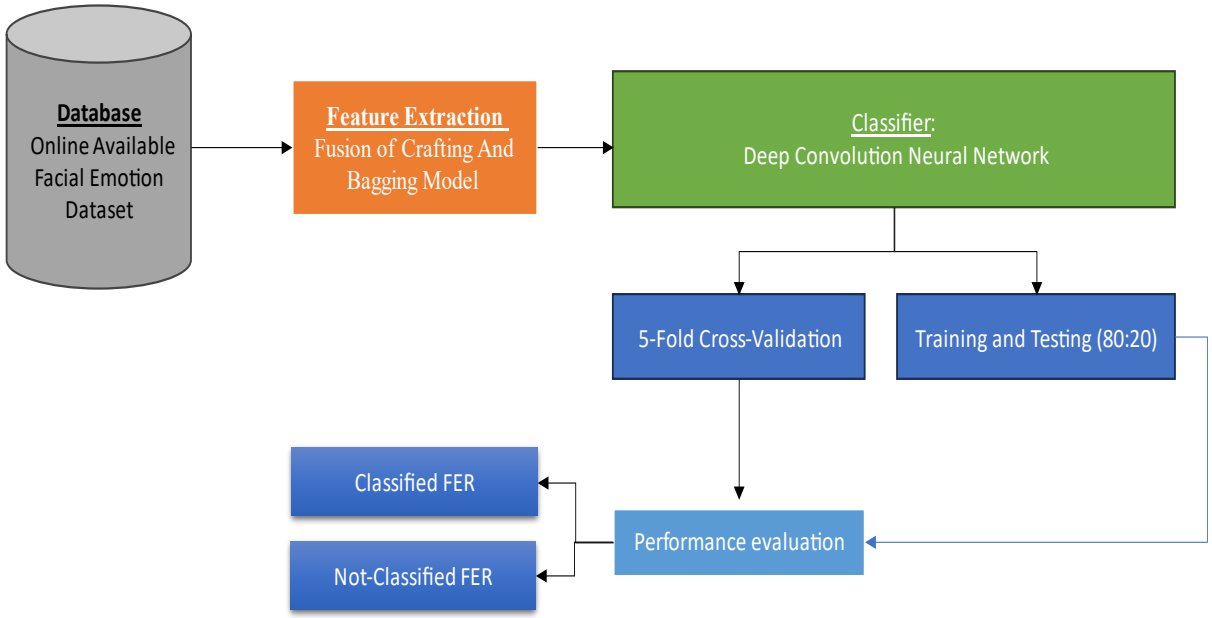

Fig 2. Overall Block Diagram of the Research Work.

The FER process includes the following steps: (1) Arranging the faces in the specific area, (2) Collecting the features of the face from the identified face area, (3) Evaluating the structure or variations encountered in the FE, such that it is classified into FE-interpretative sections like facial muscle movements such as frown or smile, emotional sections such as happiness or anger and attitude sections like ambivalence. Some original FE images have diverse sizes, complex backgrounds, shades, and other parameters. Some set of sequential image pre-processing is performed to provide FE as input for training purposes. Initially, the face in the image must be located, and the facial image must be cut off [16]. Then, normalize the image to a specific size. Subsequently, they equalize images to diminish illumination and other factors.

Finally, the edges of the image layer are extracted using a convolutional process. The extracted information is super-imposed on every feature to handle the structural data of the texture image. The ultimate objective of performing normalization is to make image features more familiar for processing and improve FE or segmentation [17]. **Fig 2** depicts the overall block diagram of the research work.

Normalization

The network will accept only fixed-size images whenever an image is provided as input. Therefore, normalization must be done with a certain-sized image instead of the original image. Let point (a, b) of the original image be normalized, and image mapping is provided as (a', b'). The mapping is provided as in EQU (1)

$$\begin{bmatrix} a' \\ b' \\ 1 \end{bmatrix} = \begin{bmatrix} s_a & 0 & 0 \\ 0 & s_b & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

where, s_a specifies scaling ratio in $'a'$ axis and s_b specifies the scaling ratio of the image $'b'$ axis. A bilinear interpolation procedure is required to handle the image NaN value in the scaling process. A, B, C,D are four points over pixel (a, b). Gray point values (x, y) and gray value computation of E, F are done with EQU (2) and EQU (3):

$$g(E) = (a - a_D)(g(C) - g(D) + g(D) \tag{2}$$

$$g(F) = (a - a_A)(g(B) - g(A) + g(A) \tag{3}$$

where a_A and a_D are A and D abscissa point correspondingly. Gray scale computations of (a, b) are provided as in EQU (4)

$$g(a, b) = (a - a_D)(g(F) - g(E) + g(E) \tag{4}$$

where a_D specifies CD ordinate points. By image normalization, the scaling image is 128×128 in size.

General Mapping with Normalization

Here, normalized facial geometry mapping is specified in m×n×3 matrices as in EQU (5):

$$I_g = [p_{ij}(x, y)]m \times n = [p_{ijk}]m \times n\{x, y\} \tag{5}$$

where, $[p_{ij}(x, y)]m \times n = (p_{ijx,}p_{ijy})\mathrm{T}, (1 \leq \mathrm{i} \leq \mathrm{m}, 1 \leq \mathrm{j} \leq \mathrm{n})\, \mathrm{i}, \mathrm{j} \in \mathrm{Z}$ specifies 2-D point p_{ij}. Unit normalization matrix is provided in EQU (6):

$$I_n = [n(p_{ij}(x, y))]\mathrm{m} \times \mathrm{n} = [n_{ijk}]m \times n \times \{x, y\} \tag{6}$$

where, $[n(p_{ij}(x, y))]\mathrm{m} \times \mathrm{n} = (n_{ijx}, n_{ijy}, n_{ijz})\mathrm{T}, (1 \leq \mathrm{i} <= \mathrm{m}, 1 \leq \mathrm{j} \leq \mathrm{n}, \mathrm{i}, \mathrm{j} \in \mathrm{Z})$ is p_{ij} unit normalized vector. Here, the local plane fitting approach is used to evaluate I_n. For all points $p_{ij} \in I_g$, corresponding normalized vector $n(p_{ij})$ is evaluated as vectors of the fitting plane as in EQU (7):

$$S_{ij}: n_{ijx}q_{ijx} + n_{ijy}q_{ijy} = d \tag{7}$$

where $(q_{ijx}, q_{ijy})\ T$ is the local neighbourhood point of $p_{ij,}$ and $d = n_{ijx}p_{ijx} + n_{ijy}p_{ijy}$. Here, the 5×5 window is cast off. To make it easier, every component is specified as m×n matrix as in EQU (8):

$$I_n = \begin{cases} I_n^x = [n_{ij}^x]m * n \\ I_n^y = [n_{ij}]m * n \end{cases} \tag{8}$$

where $\left\| \left(n_{ij}^x, n_{ij}^x\right)^T \right\|_2 = 1.$

Mapping Curve with Normalization

The local cubic fitting approach is cast off for curve normalization to compute curvatures. This approach considers surface local geometry approximated by patching [18-19]. The nominal solution for fitting uses normal vectors and 2D coordinates of neighboring points $p_{ij} \in I_g$ to be evaluated are used. This is provided in EQU (9) to EQU (11).

$$Z(\mathrm{x}, \mathrm{y}) = \frac{a}{2}x^2 + bxy + \frac{c}{2}y^2 + dx^3 + ex^2y + fxy^2 + gy^3 \tag{9}$$

$$Z_x = ax + by + 3dx^2 + 2exy + fy^2 \tag{10}$$

$$Z_y = bx + cy + 3gy^2 + 2fxy + ex^2 \tag{11}$$

These are solved with shape operator $'S'$, and least squares regression is evaluated as in EQU (12):

$$S = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{12}$$

However, $'S'$ eigen values of principle curvatures k_1 and k_2 at point $p_{ij} \in I_g$. Normalization (index value) is depicted as in EQU (13):

$$\frac{1}{2} - \frac{1}{\pi}\tan(curve)\left(\frac{k_1+k_2}{k_1-k_2}\right) \tag{13}$$

Therefore, 2D geometric of facial attributes of featured scans with expression is provided above.

Equalization

In the original acquisition procedure, it is simpler to be influenced by shadows, illumination, and parameters that lead to collected images with uneven shade and light distribution, leading to the complexity of FE [20]. Henceforth, it is essential to normalize grey-level images to improve image contradiction. Here, the Histogram Equalization approach is cast off to process images. The preliminary concept behind the histogram transformation of the original image is the uniform distribution. If gray leveling is G, size is $M \times N$, the number of pixels in r_i gray level is E, gray level probability is depicted as in EQU (14):

$$p_r(r_i) = \frac{n_i}{M*N}, \qquad i = 0,1,..,L-1 \tag{14}$$

Following this, the cumulative distribution function is computed with EQU (15):

$$T(r_i) = \sum_{j=0}^{i} p_r(r_j),\ i = 0,1,\ldots.,\mathrm{L}-1 \tag{15}$$

At last, the image histogram is computed with the following mapping EQU (16):

$$e_j = INT\left[\left(e_{max} - e_{min}\right)_{j=0,1,\ldots,\mathrm{L}-1} + e_{min} + 0.5\right] \tag{16}$$

when image histograms seem uniform, image entropy is higher, and image contrast is more considerable. However, gray level equalization is considered a uniform image histogram distribution that improves the contrast image and makes it clear and finest for facial FE.

Crafting with Bagging

A visualization bagging model is anticipated for FER, which is partitioned into two aspects: one is for testing and training. In the training phase, feature specification is done by hauling out dense SIFT descriptors from training and quantizing descriptors to visualize words with k −means clustering. These are stored in randomized forests to diminish search costs. After constructing vocabulary, testing and training pipelines are equivalent. For all images in training/testing sets, visual word occurrence is recorded in the binary feature vector. Visual bagging models depicted are eliminated with spatial relationships between visual words; however, it can attain superior performance by spatial information. Spatial representation is attained by partitioning images into bins and evaluating binary vectors in all bins. Final representation is superior, as most features, like muscle contraction at the eye corner, are viable in certain face regions.

Fusing Learning with Features

Here, the crafting and learning model is merged by concatenating respective features before initiating the classification process. Remove the SoftMax layer and consider activation mapping as feature vectors relating to the image provided as network input to haul out features from fine-tuned or fine-tuned CNN. Feature vectors are normalized with $L2$ norm. Visualizing bagging is employed as a crafting feature. These features are normalized with $L2$ norm.

Local Vs. Global Learning

Local learning techniques try to regulate training system performance to train set features in every input space area. It works explicitly as (1) selecting training samples located in provided test samples, (2) training the classifier with fewer samples, and (3) applying the classifier to recognize the class label of testing samples. Similarly, CNN uses a discriminative model dependent on training samples. It is a binary classifier that attempts to determine weighted vectors and biasing terms that demonstrate a hyperplane that separates feature vectors maximally with training samples to 2 classes. To continue this, a classifier with Multi-class FER is employed. For instance, $k-NN$ model is considered local learning. Generally, $k-NN$ model is the simplest learning formulation, as the discriminant function is constant. However, local learning is employed in all classifiers. Here, CNN is utilized for local learning.

Describing the classifier (*i.e.,*) the distance measure is vital to demonstrate the neighborhood vicinity of test samples. A standardized model uses linear classifiers globally to provide a linear discriminative function. This functionality does not remain as linear in the learning process, as prediction is provided by another classifier trained for test samples. However, the discriminative function will not be provided without any test samples. However, this model is used to repair certain classifier restrictions. Also, there are some beneficial parts to standard learning. Initially, it partitions classification crises into sub-problems. Subsequently, selecting samples diminishes sample variety in the training set.

Deep Convolutional Neural Network

Fig 3 demonstrates the pipelined architecture of the anticipated D-CNN for 2D- FER. With the provided set of pre-processed textured facial images with diverse expressions, every expression deals with six kinds of 2D attributed facial mapping geometry mapping, 3 normal element mapping (normalized vectors), texture mapping, and normalized curvature mapping. Therefore, with these facial attributes, the mapping of every textured 2D scan is merged into subsets of FE (Convolutional, ReLU, and pooling layer repetitions) with sharp factors that result in numerous Multi-channel featured maps. Featured mapping is provided into a fusion network (comprising fusion layers (two features) and re-shape) that leads to a superior concentration of facial representation (fused deep feature). Lastly, the SoftMax layer and loss are considered for expression recognition and network training.

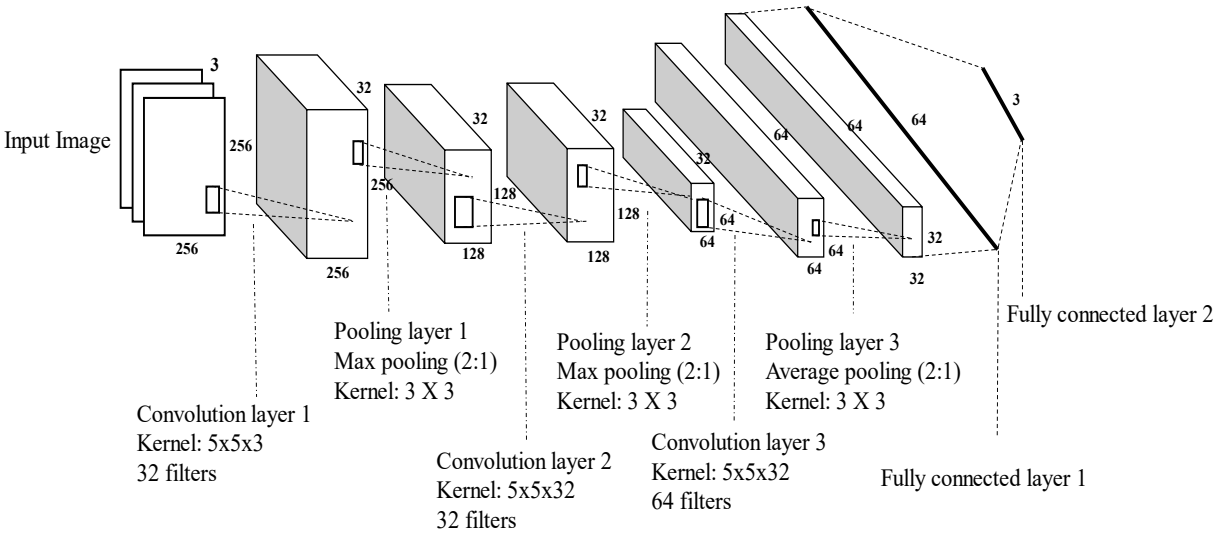

Fig 3. D-CNN model with 3 Convolutional Layers (CL), 3 RLU layers, 3 MPL layers with a Fully Connected (FC) and SoftMax layer.

For training D-CNN, assume that there is an extremely restricted amount of textured 2D facial images by labelling expression, and the FE network is initialized with the traditional CL of the pre-trained deep model. This pre-trained DL model is validated to provide superior generalization capability of generic recognition tasks. The featured network is initialized randomly, and the entire network is trained with Stochastic Gradient Descent (SGD), BP algorithm, and SoftMax loss function procedure.

For the D-CNN testing process, facial attributed mapping of every textured 2D scan is fed towards fusing feature and FE, producing superior concentration towards facial representation (deeply fused feature). Every SoftMax layer transforms This deeply fused feature into a dimensionality-based probability vector. The FER label is conducted using training CNN classifier with deeply fused features (*i.e.,*) D-CNN or openly conducting SoftMax-based prediction over dimensionality vector probability (CNN-SoftMax).

The CL is the CNN model with value sharing and local connection characteristics. CNN employs local learning in the CL of the proposed CNN architecture. This is to make the training process more robust and to improve accuracy. Some approaches use a convolution core to compute the upper layer by sliding windows. Input and convolution filters are convoluted to offer a mapping layer. Then, mapping is processed, comprising weighting and biasing operations. Mapping is attained via the activation function. Afterward, the convolution operation is performed on the new layer to produce the next layer. Following similar operations, subsequent layers are attained. Finally, features are mapped and connected to a feature vector set that is transformed into a CNN classifier for training. Usually, the CL produces computation expression as in EQU (17):

$$y_j^i = \theta\left(\sum_{i=1}^{N_j^{l-1}} w_{i,j} \otimes x_i^{l-1} + b_j^l\right), \mathrm{j} = 1,2,\ldots,\mathrm{M} \tag{17}$$

Here, $'l'$ is the present CL, $l-1$ is the layer before the new layer, $W_{i,j}$ is $j\ th$ feature graph of the convolutional kernel of the present layer and $'i'$ feature realization of the previous layer. y_j^l is j^{th} feature realization of the present layer, b_j^l is the bias of the present layer. x_i^{l-1} is feature realization of the previous layer. In analysis, $b_j^l = 0$ facilitates the network to train and diminish learning factors quickly. $'M'$ is the total realization of features at the present layer. θ (.) is the activation function; N_j^{l-1} is connected to the present layer, modified Linear Unit (ReLU) function indeed of generally utilized tangent function as ReLU uses sparse function as in EQU (18):

$$\theta\,(x) \;=\; \mathrm{Max}(0, x) \tag{18}$$

It is validated that the network is trained using the ReLU activation function with nominal sparsity. Similarly, it resolves gradient disappearance during adjustment with backpropagation factors and fastens network convergence. Features hauled out during convolutional functions are openly cast off to train classifiers. A down-sampling function is anticipated after the convolution operation to diminish these factors. The baseline for downsampling is the pixel occurrence at a continuous image range with similar global learning (correlation); therefore, features of diverse locations are counted and aggregated. For instance, compute the maximum and average values of some features of facial images. The statistical dimensionality reduction approach diminishes total parameters and eliminates fitting. However, it acquires image scaling invariance.

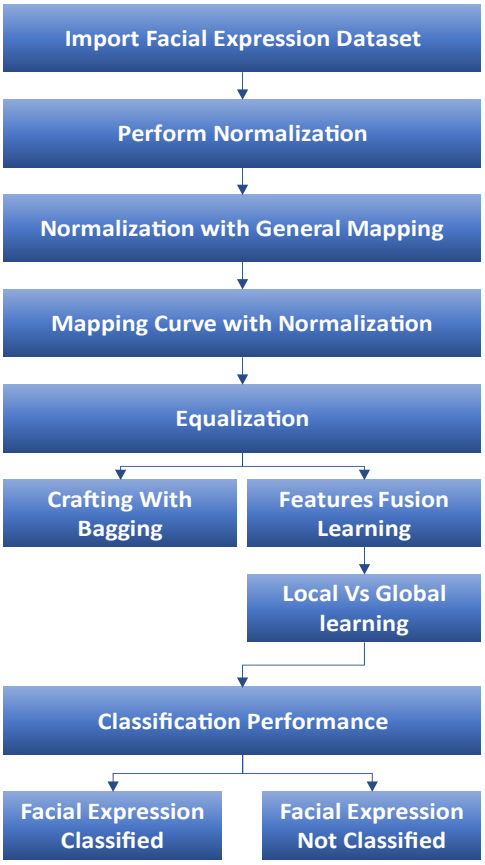

Fig 4. Flow Diagram of D-CNN Model.

Generally, the CL convolutes the input image pixel with the core region; every neuron represents the local learning field. Therefore, the convolution operation's Feature Mapping (FM) size is provided. With numerous convolutional kernels, several FM is acquired. Therefore, diverse local FE features are hauled out. The CL utilizes kernels and convolutes the output characteristics of these layers. Some set of features are attained. In successive layers, kernels are utilized to convolute mapping of pooling layer output, and features are successively attained. Every FM size is evaluated.

Then, the sharing approach is measured using the statistical characteristics of the provided image. Therefore, similar kernels haul out features from all probable image positions. Therefore, it is not suitable to utilize kernel for feature learning. Henceforth, numerous kernels are cast off for appropriate training of CNN to enhance FM. With weight sharing, data is extracted, but factors needed for training are still diminished. The generalization comparability of CNN is improved in this manner. Features hauled out from these functions are used for training classifiers. However, it encounters enormous computational factors. To reduce these facts, down-sampling functions are anticipated after the convolutional function. Pixels with similar image ranges possess similar functions; therefore, features of diverse locations are counted. For instance, a maximal or average value is computed with specific features. The statistical dimensionality reduction approach not only diminishes parameter count but eliminates fitting. Thus, it makes image scaling invariance. **Fig 4** depicts the flow of the anticipated D-CNN.

Next is the pooling function; it is cast off to diminish dimensionality. A window size of 2×2 may reduce the dimensionality of the pooling layer in half. However, there is no direct dimensionality reduction in training factors. Reduction leads to computational complexity in convolutional functionality that is reduced significantly, which enhances training speed. Generally, the generic form of pooling layer along with filters is 2×2, applied at downsamples of depth slices in input. This also discards 75% of activation, where this work performs every max operation with four numbers. If

the SoftMax classifier is trained directly with learning features, it inevitably acquires a dimensionality disaster problem. To resolve this crisis, the pooling layer is designed after the CL to diminish feature dimensionality. Downsampling will not modify total FM but diminishes FM outputs. It diminishes translation sensitivity, rotation, scaling, and other transformations if the sampling window size is $n * n$, after downsampling, the feature size of the original FM.

$$y_j^l = \theta\left(\beta_j^l down\left(y_j^{l-1}\right) + b_j^l\right) \tag{19}$$

From above EQU (19), y_j^l and y_j^{l-1} is FM of the first and present layers specifically. $down$ (.) is downsampling function; b_j^l and β_j^l is the additive and multiplicative bias of FM of the present layer. In this experiment, $\beta_j^l = 1$, $\theta(.)$, b_j^l is utilized as identical activation functions. After sharing weights, the number of training factors is diminished drastically. However, feature dimensionality is not reduced. This leads to two new crises. Initially, if the dimensions of features are enormous, the number of training parameters produced using full connection is huge. Next, the execution time is too long. The complete connection layer follows the pooling layer. It is a 1-D array where the pooling layer is a 2D array. Initially, a 2D array is related to features with a 1D array. Here, 128 1D arrays are a series of connected feature vectors. Every neuron output is provided as in EQU (20):

$$h_{w,b}(x) = \theta(w^T x + b) \tag{20}$$

where $h_{w,b}(x)$ is neuron output, $'x'$ is eigen vector input vector, $'w'$ is weighted vector, $'b'$ is bias, where $b = 0$ is activation function. $\theta(.)$ is an activation function where ReLU is cast off in these experiments. The total amount of neurons influences the network's fitting capability and training speed. Experimental outcomes demonstrate that the outcome is superior when the neuron count is 250.

Algorithm 1. D-CNN for facial ER of Students
Input: Training image: X; label: y
Ensure: D-CNN: Final Feature: expressions recognized or not
Step 1. Initialize parameters.
Step 2. Repeat
Step 3. Forwarding procedure
Step 4. Import Input Image
Step 5. Inject Normalization with EQU (21)

$$\begin{bmatrix} a' \\ b' \\ 1 \end{bmatrix} = \begin{bmatrix} s_a & 0 & 0 \\ 0 & s_b & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{21}$$

Step 6. Determine general or curvature-based normalization using EQU (22) to EQU (23):

$$I_n = \begin{cases} I_n^x = \left[m_{ij}^x\right] m * n \\ I_n^y = \left[n_{ij}\right] m * n \end{cases} \tag{22}$$

$$\frac{1}{2} - \frac{1}{\pi} \tan\left(curve\right) \left(\frac{k_1 + k_2}{k_1 - k_2}\right) \tag{23}$$

Step 7. Perform Equalization with EQU (24)

$$e_j = INT\left[\left(e_{max} - e_{min}\right)_{j=0,1,\ldots,\mathrm{L}-1} + e_{min} + 0.5\right] \tag{24}$$

Step 8. Offload some processing tasks to edge devices or servers closer to the data source, reducing latency and enhancing real-time capabilities.
Step 9. Compute global or local learning processes with visualization bagging and crafted features.
Step 10. Classify expression based on dimensionality reduction with FM
Step 11. Fit D-CNN label based on forwarding and backpropagation bias:
Step 12. Output of CNN neuron layer.
Step 13. Compute the accuracy of the prediction rate, either classified or not

Finally, the SoftMax layer is the final layer of the classifier. It is a Multi-output classifier. When samples are provided as input, output values lie between 0 and 1, which specifies the input sample probability that belongs to the class label. However, the neuron with the highest output value is chosen as the classification output.

Training with D-CNN

Here, D-CNN is provided with an essential optimization process and weighted network updation. Initial weight approximation has the highest influence on weight initialization updation. The generally used approach comprises uniform distribution, constant initialization, and Gaussian distribution. CNN specifically executes mapping association among output and input. Before commencing training, network ownership should be done with diverse random numbers. CNN training is partitioned into two stages:

Backpropagation

It is considered an error propagation stage. Compute error among y' and y as error among class label and SoftMax layer for the provided sample and adjust weight sharing by reducing mean square error.

Forward Propagation

samples are hauled out from the training set. Its categories label $'y'$ whose element specifies the probability that partitions diverse categories. $'x'$ is D-CNN network input. The upper layer output is the current layer's input. Output is computed with the activation function. Finally, the SoftMax layer output is attained.

IV. RESULT AND DISCUSSION

This section discusses the numerical results and the discussions related to the results of the proposed D-CNN model. The simulation was done in the MATLAB 2018a environment, and the system was configured with a 64-bit Operating System, 8 GB RAM, and Intel Core I5 processor, respectively. Here, two experiments are modelled to validate probability model performance. Initial work examines the performance of the anticipated algorithm and validates D-CNN training time, which is lower than the conventional CNN model. Facial ER data were acquired from the expression dataset, as shown in **Table 1**. This dataset comprises training images and testing images. It has 48×48 grayscale images. Faces are in the middle of every image. Henceforth, with experimentation, data are directly provided as inputs to the network for training purposes without any pre-processing.

Training and testing are done with an 80:20 ratio, and a 5-fold CV is done here. Generally, in the k-fold CV, all entries are from the original dataset. This is utilized for both validation and training. In general, *k* is 10; however, it is not mandatory to be strict with 10; $'k'$ can be any value. In the proposed method, $'k'$ is 5. The functionality of the model evaluates D-CNN with k-fold cross-validation. The samples are divided into training, testing, and validation phases. During CV, only 10% of the samples are used for testing. The slices are rotated and cropped in 900, 1800, and 2700) angle degrees to improve the testing and training process. The nodules are vertically and horizontally flipped. Various control measures are modelled to validate the significance of the D-CNN performance. Some conditions must be maintained for certain factors during experimentation, and diverse effects are evaluated.

Performance Metrics

The performance of the proposed D-CNN is evaluated with metrics like accuracy, precision, sensitivity, specificity, F-measure, recall, MCC, ROC, and Confusion Matrix (CM). These are evaluated with True Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP).

Accuracy is depicted as the overall measure of classification efficiency, EQU (25):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{25}$$

Sensitivity is depicted as a classifier measure to identify positive class patterns. It is expressed as in EQU (26):

$$Sensitivity = \frac{TP}{TP+FN} \tag{26}$$

Specificity is depicted to measure classifier competency to identify negative class patterns as in EQU (27):

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{27}$$

Mathew's Correlation Coefficient (MCC) is measured as the classification rate in binary class problems, ranging from -1 to +1. Here, -1 specifies the mistake or error, and +1 specifies the appropriate label. However, '0' relies on random prediction, EQU (28):

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{28}$$

Dataset Description

The FE dataset is composed of 48×48-pixel gray-scale face images. The faces are registered automatically; thus, the registered faces are positioned in the centralized region and occupy the same amount of storage space for all images. The target of this dataset is to classify the face based on facial emotion-based expressions under the given categories. **Table 1** depicts the FER-2013-based FE categorization.

Table 1. FE Categorization

Numeric Code	Pixel Column
0	Angry
1	Disgust
2	Fear
3	Happy
4	Sad
5	Surprise
6	Neutral

The *train.csv* model comprises two columns (*i.e.,*) emotion and pixels, where the emotion columns are numerical codes ranging from 0 to 1. It includes the emotion in the image. The pixel column is composed of strings for all images. The string content is space-separated pixel values (row). Similarly, the *test.csv* comprises pixel columns, and the task is to identify the emotion column. The training set is composed of 28,709 samples. A testing set (public) is utilized for the leaderboard composed of 3,589 samples. The testing set (final) is adopted to determine the competition winner. It is composed of 3,589 samples. Aaron Courviller and Pierre-Luc Carrier prepared the FER 2013 dataset as the initial part of their research work. The authors gave the samples to the workshop organizers with the primary version of the constructed dataset for research purposes. **Table 2** depicts the number of data (samples).

Table 2. Number of Data in the FER 2013 Dataset

Classification	Validation Data		Training Data	Total Dataset
	Public	Private		
Angry	497	491	3995	4958
Disgust	56	55	436	547
Fear	496	528	4097	5121
Happy	895	879	7215	8989
Sad	653	594	4830	6077
Surprise	415	416	3171	4002
Neutral	607	626	4965	6198
	3589	3589	28709	35887

Performance Evaluation

The performance of the proposed DCNN is compared with the existing methods, namely Support Vector Machine (SVM), Hidden Markov Model (HMM), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). **Fig 5** shows sample input image. **Table 3** shows comparison of performance.

Table 3. Comparison of Performance

Algorithm	Accuracy	Sensitivity	Specificity	MCC
SVM	89	50.3	87.6	40.5
HMM	89.2	48.4	87.1	38.3
CNN	91	49.9	89	48.7
LSTM	86.8	49.2	84.4	46.9
DCNN	99.5	50.5	98.3	79.2

Fig 5. Sample Input Image.

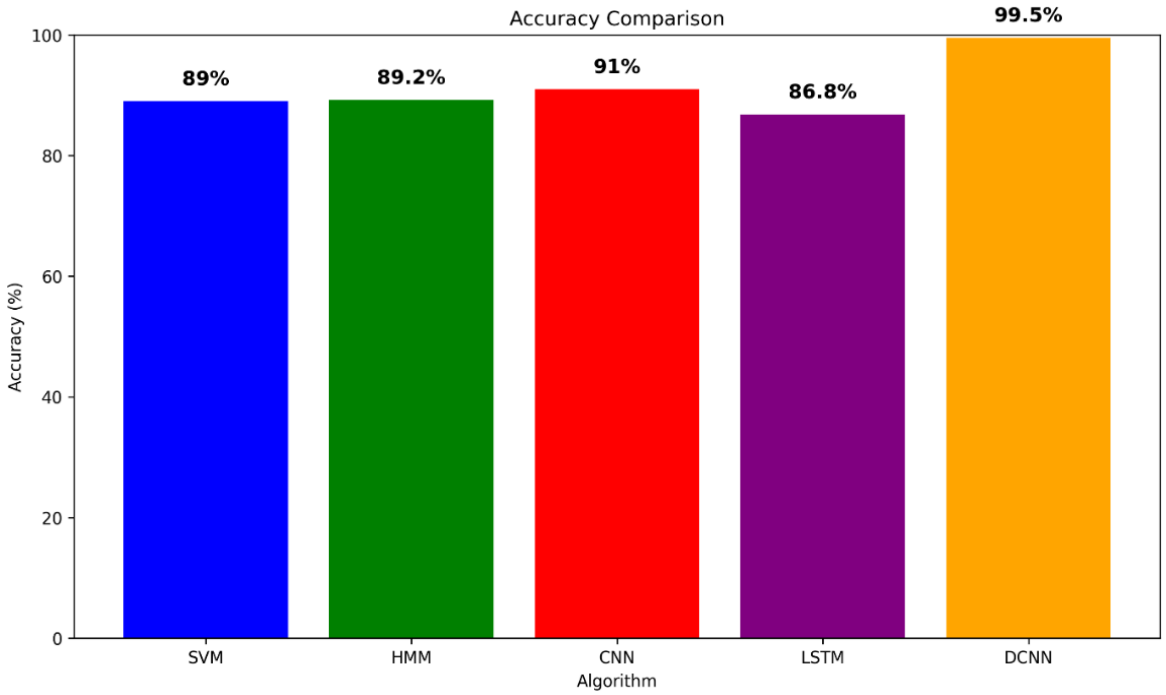

Fig 6. Comparison of Accuracy.

In comparing algorithm performance, SVM and HMM achieved similar accuracies at around 89%, while CNN demonstrated improved accuracy at 91%. LSTM performed slightly lower at 86.8%, but DCNN outperformed all with an accuracy of 99.5%. Selecting a suitable algorithm is vital for effective ER systems. **Fig 6** shows comparison of accuracy.

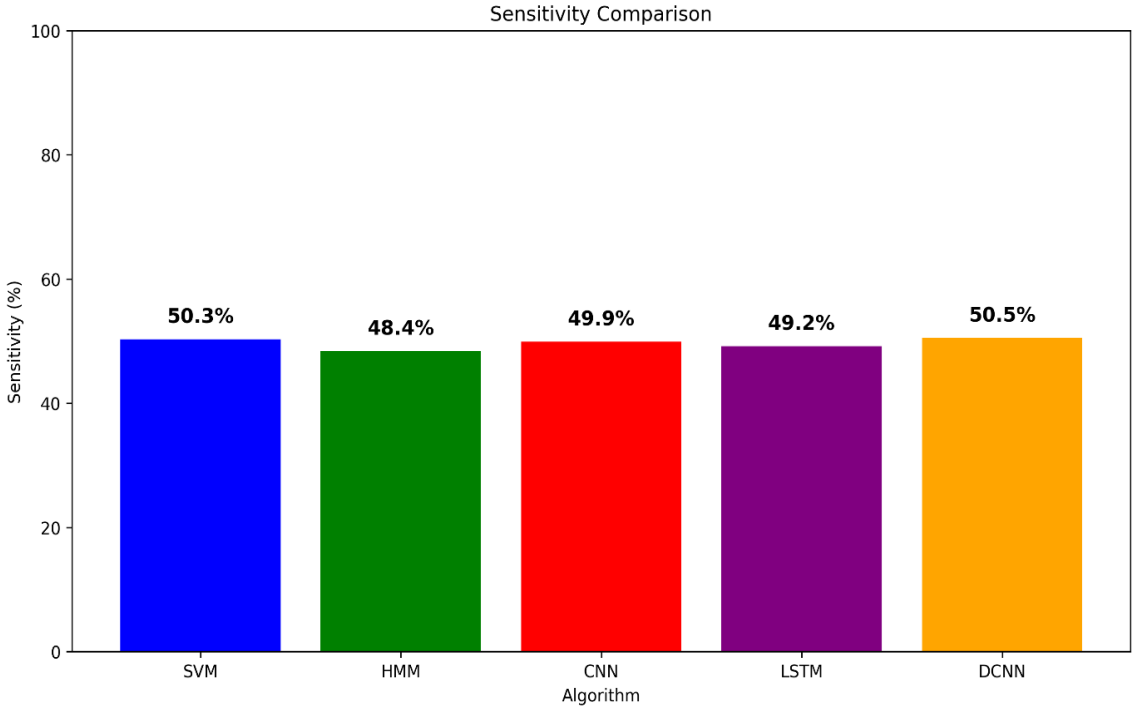

Fig 7. Comparison of Sensitivity.

Comparing the algorithm sensitivities, SVM, CNN, and LSTM all exhibited similar performance, ranging from 48.4% to 50.5%. These results indicate a relatively consistent ability to correctly identify positive cases across these models. However, HMM displayed a slightly lower sensitivity at 48.4%. **Fig 7** shows comparison of sensitivity.

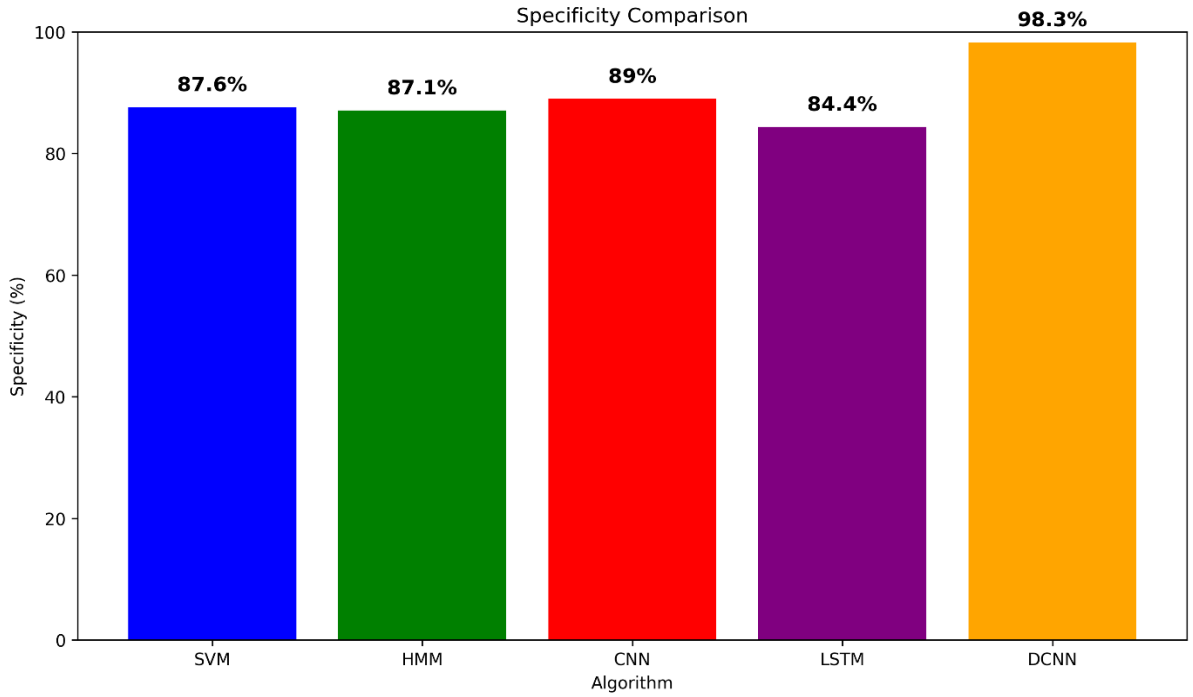

Fig 8. Comparison of Specificity.

In the comparative analysis of algorithm specificity, DCNN stands out with the highest score of 98.3%, highlighting its exceptional ability to identify negative cases correctly. CNN follows closely at 89%, indicating impressive performance. SVM and HMM exhibit similar but lower specificities at 87.6% and 87.1%, respectively, while LSTM lags at 84.4%. **Fig 8** shows comparison of specificity.

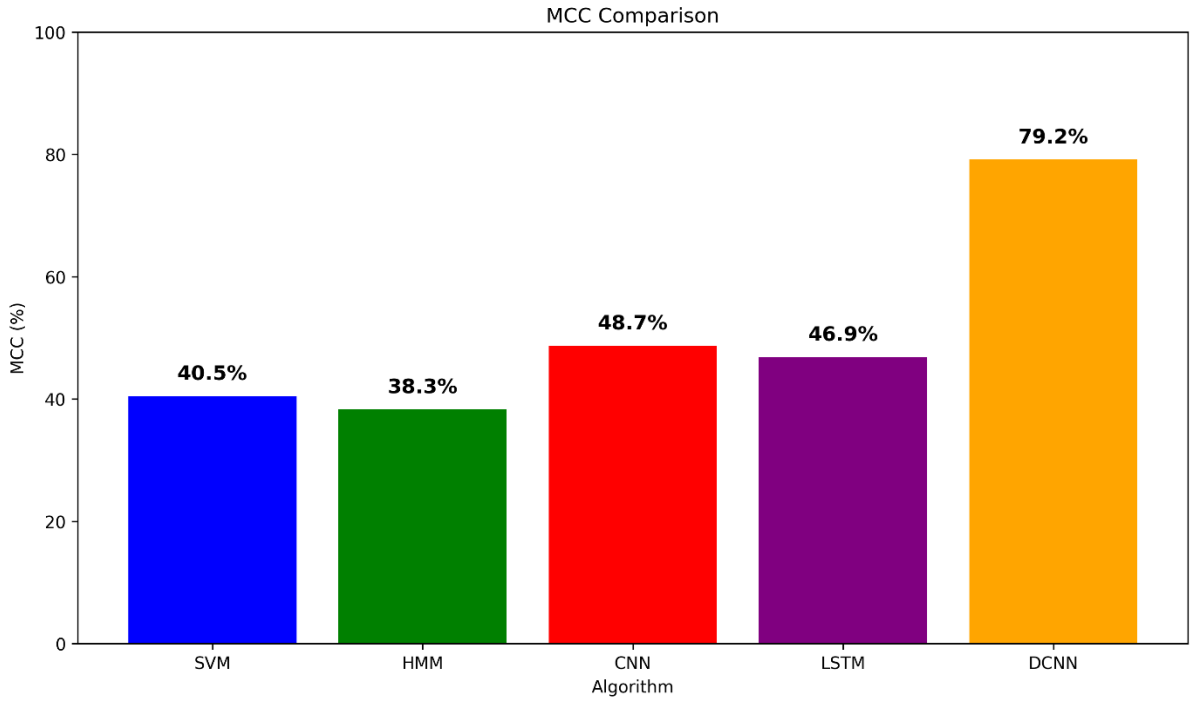

Fig 9. Comparison of MCC.

Comparing the MCC scores, DCNN significantly outperforms other algorithms with an impressive 79.2, demonstrating its exceptional balance between TP and TN. CNN follows closely with 48.7, indicating overall solid performance. SVM and HMM show lower MCC values at 40.5 and 38.3, respectively, while LSTM is at 46.9. F**ig 9** shows comparison of MCC.

V. CONCLUSION AND FUTURE WORK

Deep Learning (DL) algorithm-based research on emotion identification and predicting college students' Mental Health (MH) has enormous potential. It addresses significant difficulties in the early detection of mental health problems and individualized support, potentially lowering dropout rates and enhancing well-being. This research topic is crucial and timely due to the ethical ramifications, data-driven insights, and possible societal influence. Utilizing DL algorithms for MH support on college campuses is essential in advancing students' general well-being and founding a healthier learning environment as a technological development. The examination of ER algorithms reveals significant performance variances. As the best algorithm, DCNN is superior in MCC scores and perfectly balances TP and TN. While SVM and HMM perform only somewhat, CNN performs well. Between these two extremes is where LSTM lies. The algorithm chosen should align with a given application's priorities and requirements; DCNN and CNN are good options for assignments needing high MCC scores. Proactively supporting college students will take a quantum leap with the combination of IoT-enabled Edge Computing (EC) and DL algorithms for emotion identification and MHP. This technology can completely change how educational institutions and healthcare professionals handle MH and emotional well-being academically and improve the existing support system. The system should be improved, ethical issues should be addressed, and the system's use should be expanded to assist a larger student population.

Optimized DL algorithms should be used in future studies on Emotion Recognition (EM) and Mental Health Prediction (MHP) for college students. These include developing real-time monitoring systems for prompt assistance, incorporating multi-modal data sources to increase accuracy, and longitudinal research to monitor changes in MH. Concentrating on understandable AI and individualized treatments while considering ethical considerations and cultural sensitivity is crucial. The active participation of users in the design process and collaboration with MH specialists are essential, as is the evaluation of long-term effects. Additionally, testing on numerous datasets, investigating hybrid models, and applying user-centered design principles can help create AI-based MH care systems that are more efficient and moral.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests

Reference

[1]. J. H. Cheong, E. Jolly, T. Xie, S. Byrne, M. Kenney, and L. J. Chang, "Py-Feat: Python Facial Expression Analysis Toolbox," Affective Science, vol. 4, no. 4, pp. 781–796, Aug. 2023, doi: 10.1007/s42761-023-00191-4.

[2]. H. H. Nguyen, V. T. Huynh and S. H. Kim, "An Ensemble Approach for Facial Expression Analysis in Video," 2022, arXiv preprint arXiv:2203.12891.

[3]. Online accessible: https://www.kaggle.com/deadskull7/fer2013

[4]. F. Principi, S. Berretti, C. Ferrari, N. Otberdout, M. Daoudi, and A. Del Bimbo, "The Florence 4D Facial Expression Dataset," 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), vol. 20, pp. 1–6, Jan. 2023, doi: 10.1109/fg57933.2023.10042606.

[5]. F. Xue, Z. Tan, Y. Zhu, Z. Ma, and G. Guo, "Coarse-to-Fine Cascaded Networks with Smooth Predicting for Video Facial Expression Recognition," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2022, doi: 10.1109/cvprw56347.2022.00269.

[6]. K. Wolf, "Measuring facial expression of emotion," Dialogues in Clinical Neuroscience, vol. 17, no. 4, pp. 457–462, Dec. 2015, doi: 10.31887/dcns.2015.17.4/kwolf.

[7]. H. Ge, Z. Zhu, Y. Dai, B. Wang, and X. Wu, "Facial expression recognition based on deep learning," Computer Methods and Programs in Biomedicine, vol. 215, p. 106621, Mar. 2022, doi: 10.1016/j.cmpb.2022.106621.

[8]. W. Yu and H. Xu, "Co-attentive multi-task convolutional neural network for facial expression recognition," Pattern Recognition, vol. 123, p. 108401, Mar. 2022, doi: 10.1016/j.patcog.2021.108401.

[9]. H.-S. Cha and C.-H. Im, "Performance enhancement of facial electromyogram-based facial-expression recognition for social virtual reality applications using linear discriminant analysis adaptation," Virtual Reality, vol. 26, no. 1, pp. 385–398, Sep. 2021, doi: 10.1007/s10055-021-00575-6.

[10]. M. Monaro, S. Maldera, C. Scarpazza, G. Sartori, and N. Navarin, "Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models," Computers in Human Behavior, vol. 127, p. 107063, Feb. 2022, doi: 10.1016/j.chb.2021.107063.

[11]. Y. Nan, J. Ju, Q. Hua, H. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recognition," Alexandria Engineering Journal, vol. 61, no. 6, pp. 4435–4444, Jun. 2022, doi: 10.1016/j.aej.2021.09.066.

[12]. N. Shabbir and R. K. Rout, "Variation of deep features analysis for facial expression recognition system," Multimedia Tools and Applications, vol. 82, no. 8, pp. 11507–11522, Nov. 2022, doi: 10.1007/s11042-022-14054-w.

[13]. S. Li et al., "Facial Expression Recognition In-the-Wild with Deep Pre-trained Models," Computer Vision – ECCV 2022 Workshops, pp. 181–190, 2023, doi: 10.1007/978-3-031-25075-0_14.

[14]. C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, "Impact of Deep Learning Approaches on Facial Expression Recognition in Healthcare Industries," IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5619–5627, Aug. 2022, doi: 10.1109/tii.2022.3141400.
[15]. A. R. Khan, "Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges," Information, vol. 13, no. 6, p. 268, May 2022, doi: 10.3390/info13060268.
[16]. H. Sikkandar and R. Thiyagarajan, "Deep learning based facial expression recognition using improved Cat Swarm Optimization," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 2, pp. 3037–3053, Aug. 2020, doi: 10.1007/s12652-020-02463-4.
[17]. I. Gogić, M. Manhart, I. S. Pandžić, and J. Ahlberg, "Fast facial expression recognition using local binary features and shallow neural networks," The Visual Computer, vol. 36, no. 1, pp. 97–112, Aug. 2018, doi: 10.1007/s00371-018-1585-8.
[18]. P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," Applied Intelligence, vol. 52, no. 4, pp. 4543–4581, Jul. 2021, doi: 10.1007/s10489-021-02550-9.
[19]. K. Chengeta and S. Viriri, "A Review of Local, Holistic and Deep Learning Approaches in Facial Expressions Recognition," 2019 Conference on Information Communications Technology and Society (ICTAS), Mar. 2019, doi: 10.1109/ictas.2019.8703521.
[20]. G. Li and C. Le, "Hybrid Binary Bat Algorithm with Cross-Entropy Method for Feature Selection," 2019 4th International Conference on Control and Robotics Engineering (ICCRE), Apr. 2019, doi: 10.1109/iccre.2019.8724270.